



Effects of prosody in processing speaker commitment in French

Caterina Petrone¹, Alessandra Lonobile², Christelle Zielinski^{1,3} & Kiwako Ito⁴

¹ Aix-Marseille Université, CNRS, LPL UMR 7039, Aix en Provence, France,

² Università degli Studi di Napoli «Federico II», Napoli, Italy

³ Brain and Language Research Institute, AMU, Aix en Provence, France

⁴ Ohio State University, Columbus OH, US

caterina.petrone@lpl-aix.fr; alessandra.lonobile@gmail.com;
christelle.zielinski@blri.fr; ito.19@osu.edu

Abstract

In French, an utterance-final fall is often associated to commitment on speaker's behalf and it is typically used in assertions. Final rises and final rise-fall-rises signal that the speaker does not commit to the proposition of the sentence. Hence, they are often used to convey incredulity. This study tested whether listeners use earlier prosodic cues as well as the final contour in the sentences to achieve a pragmatic interpretation of an utterance. Sixteen Subject-Verb-Object sentences were presented as assertions and incredulity questions. Both prenuclear (e.g., expanded pitch range) and nuclear (e.g., final boundary tone) differentiated the intentions. Twenty-two listeners matched each auditory stimulus with one of the two facial expressions, while their eye movements were monitored. For assertions, listeners looked at the congruent picture only after listening to the whole sentence. However, for incredulity questions, anticipatory fixations to the referent picture gradually increased from the beginning of the sentence. The findings suggest that the interaction between prenuclear and nuclear contours in processing speaker commitment may vary across different tunes.

Keywords: Intonation, French, Commitment, Eye-tracking, Prenuclear contour.

1. Introduction

A hot issue in intonation research is to determine how intonational meaning is created. In the British School approach the nuclear tune, that spans from the last pitch-accent to the end of the utterance, has a special status both formally and functionally. Thus, analysing tune meanings is often equated to analysing the meaning of the nucleus [1, 2]. In contrast, the Autosegmental-Metrical (AM) framework assumes that the nuclear pitch accent is merely the last accent within a specific major prosodic phrase [3]. The overall meaning of a tune results from the independent contributions of its freely combinable, morpheme-like sub-parts that include pitch accents and edge tones [4].

Despite this compositional approach to tune meaning, work within the AM theory continued to regard the nucleus – more or less implicitly – as the semantic ‘heart’ of the tune. In particular, the nucleus has been assigned a crucial role in distinguishing different degrees of speaker's commitment. Commitment refers to the speaker's attitude towards the propositional content of his utterance. For instance, while assertions are associated to commitment on speaker's behalf, incredulity questions signal the speaker's disbelief or reduced certainty [5, 6]. In many languages (including French),

assertions and incredulity questions might be syntactically the same, so that the difference in meaning may take identical syntactic forms and thus their semantic distinction may rely on prosody [6, 7]. Assertions are often marked with final falls, whereas incredulity questions may carry an expanded pitch range or a specific tune for the nucleus accent, final rises with higher boundary tone and a longer duration of the utterance-final syllable [5, 6, 8, 9, 10].

Some behavioral studies found that the prenuclear contour (i.e., the section of the intonation contour which is realized before the nucleus) can also make a separate contribution to intonational meaning. For instance, F0 declination, global pitch range or the shape of prenuclear accents signal different degrees of commitment [7]. A limitation of these studies is that they are based on offline methodologies, which are often difficult to accomplish (e.g., gating tasks) and that don't allow to track the time course of utterance interpretation.

In this paper, we explore whether a prenuclear tune leads to the anticipation of the pragmatic intent of utterances in French. French has two prosodic units: the Accentual Phrase (AP) and the Intonation Phrase (IP), and an obligatory pitch accent LH*, associated with a metrically strong syllable, marks the right edge of non-final APs [11]. Optionally, an additional phrasal LHi may mark the left edge of the AP. The occurrence of the LHi depends on many factors, such as speech rate and phonological length of the AP. A ‘prenuclear contour’ is defined as the prosodic tune composed of both pitch accents (LH*) and phrasal tones (LHi), stretch of the contour, which precedes the nuclear accent. The existing models of French prosody identifies three nuclear tunes: rises (H*H%), falls (L*L%) and rise-falls (H*L%) [11, 12]. It has been claimed that assertions are expressed with L*L%, while the H*H% is typically used for questions. Studies report that both a nuclear H* and a nuclear H+!H* are frequently used to convey incredulity [6, 13, 14].

Informal observations suggest that the F0 contour spanning from the prenuclear to the nuclear accent differentiate assertions from incredulity questions (Fig. 1: assertion (top) and incredulity question (bottom)). The incredulity questions exhibit more expanded pitch ranges for the incredulity questions exhibit more expanded pitch ranges as compared to assertions.

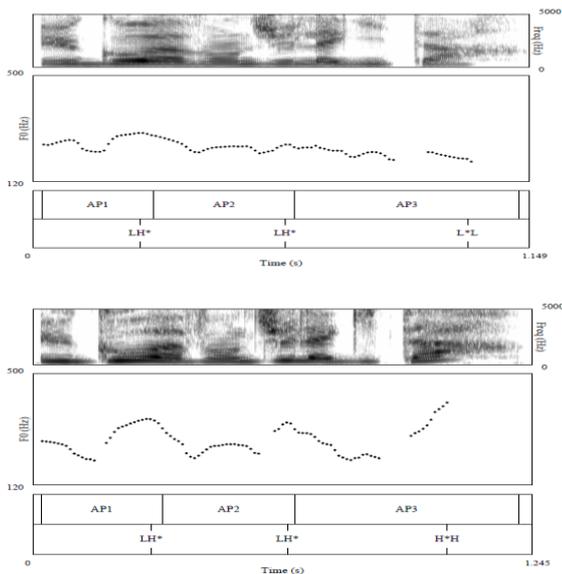


Figure 1. Two renderings of the utterance [Hugo] [a vendu] [la guitare] (“Hugo sold the guitar”), where the square brackets stand for the AP boundaries. The IP final contour is L*L% in the assertion (top) and H*H% in the incredulity question (bottom).

The present study used an eye-tracking task to test listeners’ online responses to prenuclear contour. Studies have reported an anticipatory effect of prosodic prominence [15] and the effect of boundary tones to elicit assertions vs. yes/no questions judgments [16]. However, the stimuli of [16] were limited to less typical syntactic constructions of American English (elliptical utterances only containing the nucleus). The current study presents long utterances contrasting degrees of commitment (assertions vs incredulity question) with pictures showing the corresponding facial expressions. This method is based on the previous cross-linguistic study that reports listeners’ use of facial gestures for the identification of emotions such as incredulity or surprise [17]. If the percept of speaker commitment relies solely on the nuclear information visual detection of the matched face should become reliable only after listeners heard the entire sentences. If nuclear contours lead to a percept of speaker commitment, detection of the matched face may surface before the end of the sentences.

2. Rating task for material selection

The stimuli for the eye-tracking study consisted of natural sentences produced by two French native speakers. A rating task was conducted prior to the eye-tracking study to select the stimuli.

2.1 Corpus

A male phonetician and a female professional actress from Southern France produced a set of 16 syntactically declarative sentences, containing words whose meanings were not emotionally laden. Target sentences had Subject-Verb-Object constituents, each of which syntactic constituent being 2-5 syllables long. All Subjects were a proper name, followed by a transitive verb (auxiliary and past participle) and the object NP of an article and a noun (e.g., *Hugo a vendu la guitare*, “Hugo sold the guitar”). The object NP contained the nucleus. Eight additional sentences were added as fillers and produced with four basic emotions (fear/anger; sadness/joy).

To elicit the assertion, each speaker was asked to report a given event to a fictitious interlocutor. For the incredulity question, the speaker was asked to overtly express a doubt in what his/her interlocutor has just said

Speakers silently read the contexts and then uttered the sentences aloud. They sat in front of a camera and pictures of their faces were taken during the production of the nuclear configuration such that facial distinctions among the six emotions would be maximal. Speakers were given no instructions as to what facial expression to use during sentence production. The face-to-camera distance, lighting, and background were controlled. A total, of 64 target sentences (16 sets X 2 sentence types X 2 speakers) and 64 filler sentences (8 sentence sets X 4 emotions X 2 speakers) was recorded in the anechoic room of the *Laboratoire Parole et Langage*.

Table 1. Example of contexts eliciting either the assertion (left) or the incredulity question (right). The target sentence is *Hugo a vendu la guitare* (“Hugo sold the guitar”). Contexts are translated from French.

Context for assertions.	Context for incredulity questions.
Your cousin Hugo sold his guitar. You tell it to your father:	Your friend Hugo is facing financial difficulties and he is obliged to sell some of his goods to get some money, with the exception of his beloved guitar. Once, you meet Hugo’s brother who tells you that he sold his guitar. You cannot believe that and put into question what his brother has just said:

2.2 Participants and procedure

Fifteen participants (8F/7M aged 18-35 y.o.) from Southern France, completed the rating task. They were presented with the auditory stimuli alone and they judged the degree of elicited attributes along 5-point Likert scales. For the target sentences, the scale was labeled as 1=“very certain” and 5=“very uncertain”. For fillers, the degree of each basic emotion was judged along the scale.

2.3 Results

Twenty-five 25 of 64 target sentences were well differentiated along the certainty scale, resulting in an average rating score of either <1.5 or > 4.5. The remaining 39 sentences were judged as ambiguous: their rating score was between 1.5 and 4.5. Of the 25 unambiguous sentences, 12 were recorded as assertions (10 from the F speaker, 2 from the M speaker) and 13 as incredulity questions (8 from the F speaker, 5 from the M speaker). Each sentence’s rating score was coded to be included in the eye-tracking data analysis: while unambiguous sentences, may lead to clearer anticipatory eye-movements, ambiguous sentences may lead to frequent fixations to the incorrect picture.

3. Eye-tracking study

3.1 Participants and procedure

Twenty-two native listeners (8M/14F; 18-42 y.o.) of Southern French who did not participate in the rating task took part in the eye-tracking study. Their eye movements were traced during a monitor-based dichotomous choice identification task. For each trial, participants were presented two pictures with an auditory stimulus. The pictures represented two distinct facial expressions of the same speaker. In each critical trial participants saw a picture taken toward the end of an assertion (e.g., with raised eyebrow) and another taken with an incredulity question (e.g., with furrowed eyebrow and squinted eyes) side by side, and heard either an assertion or incredulity question sentence.

Listeners were seated in front of a Tobii T120 eye tracker monitor at a distance of about 60 cm from the screen. After calibration, they were instructed to listen to the sentences and click on a facial expression that better matches the voice as quickly as possible. The auditory stimuli onset was synchronized with the appearance of the pictures on the screen. Listeners' eye locations were sampled at 60 Hz. Mouse click reaction times (RTs) were measured from the end of each auditory stimulus. The 128 trials were randomized and presented in one block. The experiment lasted for about 10 minutes. The experimental session was preceded by a short practice block with unrelated trials produced by a different speaker.

3.2 Results

Only correctly responded trials for which the mouse click took place after the sentence offset were included for the analysis. A total of 15% (with the female speaker) and 22% (with the male speaker) of eye-movement data were excluded.

Two AOIs were defined by splitting the screen in two equal parts (left and right), each including one of the two pictures. A logit mixed effects model was run, in which we tested the effects of the fixed factors REGION (Subject/Verb/Object/Mouse Click), COMMITMENT (Assertion/Incredulity Question) and SPEAKER (Male/Female) and RATING (unambiguous/ambiguous) on fixation proportions. Only pairwise interactions among fixed factors were entered in the model. Contrasts among successive levels of the factor REGION were computed, to verify the impact of the prosodic cues as the sentences unfold over time. LISTENERS and TRIALS were included as random intercepts, with additive shifts for each level of each fixed factors and their combination. An additional observational-level random intercept was added to take into account overdispersion. A backward elimination approach was adopted, starting from a complex model and deleting terms with insufficient explanatory power.

Figure 2-3 illustrate results for perceptually unambiguous and for perceptually ambiguous sentences, respectively. Each figure shows proportions of fixation on the correct picture while listening to sentences uttered as assertions (blue line) and incredulity questions (red line), averaged across trials and listeners.

Prior to calculating mean proportions, fixation data were first time-locked at 0 ms at the beginning of each sentence. Then, four temporal landmarks were annotated (beginning of the sentence, end of Subject, Verb and Object syntactic constituents, and mouse click position), thus resulting in four analysis windows or "regions": from sentence beginning to end of Subject ("Subject" region), from end of Subject to end of Verb ("Verb" region), from end of Verb to end of Object ("Object" region) and from end of

Object to Mouse click ("Mouse click" region). The whole mean proportion curve was reconstructed by joining data segments extracted from the four windows. To do that, the temporal position of the four landmarks was averaged across auditory stimuli within the same condition of commitment. This temporal realignment allowed normalizing differences in length of the syntactic constituents across target sentences (thus affecting the length of the corresponding accental phrases.)

When listening to the incredulity questions uttered by the male speaker, fixation proportions on the correct picture increased as function of REGION in a progressive fashion. Hence, fixations were higher in the Verb than in the Subject Region [$\beta = 1.73$, SE = 0.33, $z = 5.15$, $p < 0.001$], in the Object than in the Verb region [$\beta = 1.46$, SE = 0.33, $z = 4.41$, $p < 0.001$]; and in the Mouse Click than in the Object region [$\beta = 1.60$, SE = 0.33, $z = 4.76$, $p < 0.001$]. On the contrary, when listening to assertions, fixation proportions significantly increase only in the Mouse Click region. All contrasts were not significant, except the one between the Mouse Click and the Object region [$\beta = 2.52$, SE = 0.41, $z = 6.13$, $p < 0.001$].

The fixation patterns were similar for the female speaker, with the only difference that, for assertions, participants first looked at the picture representing the assertive face and next shifted gaze while listening to the auditory stimulus. Hence, proportion of fixations on the correct picture was lower in the Verb than in the Subject region [$\beta = -1.09$, SE = 0.38, $z = 2.83$, $p = 0.004$]. This might be an artifact, which arises from the fact that, in our experiment, the target picture (i.e., the picture with assertive facial expression) was always at the left position on the screen (The relative position of the pictures was not counterbalanced.) Finally, differences across RATING did not reach significance.

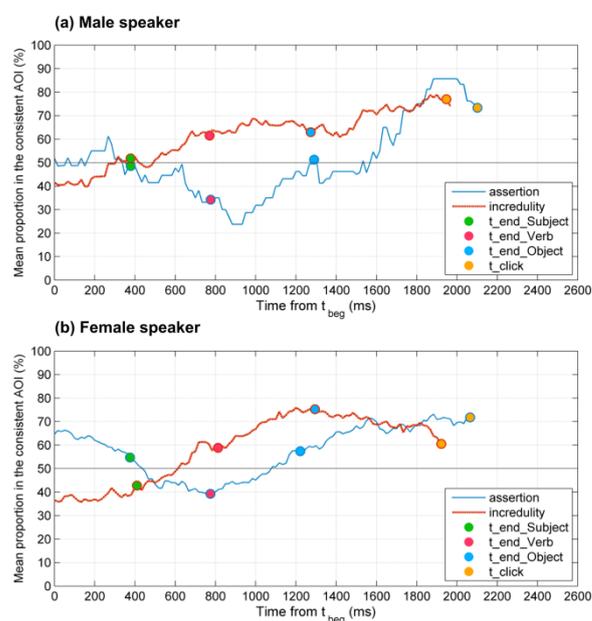


Figure 2. Mean proportion of fixations to the target pictures for sentences uttered as incredulity questions (red line) and as assertions (blue line). Data refers to perceptually unambiguous sentences. Results are pooled across all listeners. The coloured dots mark the end of the analysis windows.

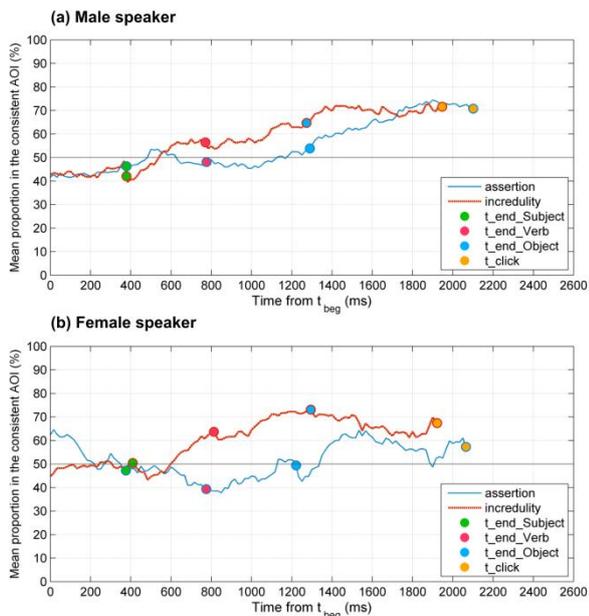


Figure 3. Mean percent of fixations and standard errors for stimuli uttered as incredulity questions (red line) and as assertions (blue line). Data refers to perceptually ambiguous sentences. Results are pooled across all listeners. The coloured dots mark the end of the analysis windows.

3. Discussion

The present results suggest that prosodic contours may lead to the immediate percept of speaker certainty in French. We focused on assertions (which convey higher speaker commitment to his/her utterance) and on incredulity questions (which convey lower speaker commitment/disbelief). After the end of the Object constituent (where the nucleus was located), fixations on the target picture exhibited a sharp increase. This is in line with a previous study in American English, which have shown immediate use of the nuclear accents and boundary tones for processing questioning vs. asserting utterances [16].

Importantly, the present data demonstrate that the pragmatic contrast is not restricted to the nucleus. When listening to the incredulity questions fixations to matched face increased by the end of the verb constituents. Hence, listeners are able to exploit the cues in the prenuclear contour and start an “anticipatory” processing for assessing speaker’s commitment. In contrast, the prenuclear contour did not seem to provide robust enough cues to evoke the “asserting” meaning. Thus, the prenuclear cues to speaker commitment are not equivalent across the utterance types for those utterances.

The ability of listeners to process incredulity questions with sole reference to the prenuclear section could result from a number of prosodic cues, such as expanded pitch range, absence of downstep, or specific phonological properties of the prenuclear contour. The current experiment does not allow for the examination of the relative contributions of various cues. Similarly, the contribution of pitch accents and edge tones for the nuclear contour cannot be evaluated separately. The two speakers produced sentences with different degrees of commitment based on specific pragmatic contexts yet they were not asked to produce a specific set of intonation contours. While the present study may have been limited by the inter-speaker variability in expressing the degrees of commitment, the results also indicate that

prosodic cues to convey similar intentions may largely vary across speakers. In addition, the lack of a RATING effect may suggest that listeners may tune to the subtle prosodic cues during online speech processing, although those cues may not weigh much for offline conscious judgments of speaker intent.

In summary, the results indicate that French listeners can make immediate use of prenuclear cues for processing speaker commitment and that the effect of nuclear and prenuclear cues may vary across different utterance types. Future studies should explore what acoustic features of the prenuclear contours provide primary cues to speaker intent under what pragmatic contexts.

4. Acknowledgements

This research is supported by the Labex *Brain and Language Research Institute*. We also thank prof. F.M. Dovetto, (University “Federico II”, Naples) that offered to the second author an Erasmus Placement grant.

5. References

- [1] M. A. K. Halliday, *An introduction to functional grammar*, London, UK: Edward Arnold, 1985.
- [2] A. Cruttenden, *Intonation*, Cambridge, MA: Cambridge University Press, 1986.
- [3] J. B. Pierrehumbert, *The Phonology and Phonetics of English intonation*, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 1980.
- [4] J. B. Pierrehumbert and J. Hirschberg, “The meaning of intonational contours in the interpretation of discourse”, in P. R. Cohen, J. Morgan, and M. E. Pollack (Eds.), *Intentions in Communication* (pp. 271–311), Cambridge: Massachusetts Institute of Technology Press, 1990.
- [5] G. Ward and J. Hirschberg, “Implicating Uncertainty: The Pragmatics of Fall-Rise Intonation”, *Language*, Vol. 61, No. 4, 747-776, 1985.
- [6] C. Portes, C. Beyssade, A. Michelas, J.-M. Marandin and M. Champagne-Lavau, “The Dialogical Dimension of Intonational Meaning: Evidence from French”, *Journal of Pragmatics*, 74, 15-29, 2014.
- [7] C. Petrone and O. Niebuhr, “On the intonation in German intonation questions: The role of the prenuclear region”, *Language and Speech* 57, 108-146, 2013.
- [8] J. B. Pierrehumbert and S. Steele, “Categories of Tonal Alignment in English”, *Phonetica* 47, 181-196, 1990.
- [9] M. Savino and M. Grice, “The role of pitch range in realizing pragmatic contrasts-The case of two questions types in Italian!”, *ICPhS XVI*, 1037-1040, 2007.
- [10] V. Crespo-Sendra, M. M. Vanrell, and P. Prieto, “Information seeking questions and incredulity questions: Gradual or categorical contrast?”, *Speech Prosody* 2010, 100164:1–4., 2010.
- [11] S-A Jun and C. Fougeron, “A phonological model of French intonation”, in A. Botinis [Ed], *Intonation: Analysis, modelling and technology*, 209-242, Boston: Kluwer Academic Publishers, 2000.
- [12] B. Post, *Tonal and phrasal structures in French intonation*, The Hague: Holland Academic Graphics, 2000.
- [13] A. Michelas, C. Portes and M. Champagne-Lavau, “Intonational contrasts encode speaker’s certainty in neutral vs incredulity questions in French”, *Proceedings of Interspeech*, France, 2013.
- [14] A. Michelas, C. Portes and M. Champagne-Lavau, “When pitch accents encode speaker commitment: evidence

from French intonation”, *Language and Speech*, published online before print June 3, 2015.

[15] K. Ito and S. R. Speer, “Anticipatory effects of intonation: Eye movements during instructed visual search”, *Journal of Memory and Language* 58, 541-573, 2008.

[16] W. F. L. Heeren, S. A. Bibyk, C. Gunlogson and M. K. Tanenhaus, “Asking or Telling – Real-time Processing of Prosodically Distinguished Questions and Statements”, *Language and Speech*: 1-28, 2015.

[17] V. Crespo Sendra, C. Kaland, M. Swerts and P. Prieto, “Perceiving incredulity: The role of intonation and facial gestures”, *Journal of Pragmatics*, 2013.